

Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User-and Marketer-Generated Content

Author(s): Khim-Yong Goh, Cheng-Suang Heng and Zhijie Lin

Source: *Information Systems Research*, Vol. 24, No. 1, Special Issue on Social Media and Business Transformation (March 2013), pp. 88-107

Published by: INFORMS

Stable URL: <https://www.jstor.org/stable/42004271>

Accessed: 01-06-2020 04:42 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Information Systems Research*

Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content

Khim-Yong Goh, Cheng-Suang Heng

School of Computing, National University of Singapore, Singapore 117417, Republic of Singapore
{gohky@comp.nus.edu.sg, hengcs@comp.nus.edu.sg}

Zhijie Lin

School of Business, Nanjing University, Nanjing 210093, People's Republic of China; and School of Computing, National University of Singapore, Singapore 117417, Republic of Singapore, mailto:zjlin@gmail.com

Despite the popular use of social media by consumers and marketers, empirical research investigating their economic values still lags. In this study, we integrate qualitative user-marketer interaction content data from a fan page brand community on Facebook and consumer transactions data to assemble a unique data set at the individual consumer level. We then quantify the impact of community contents from consumers (user-generated content, i.e., UGC) and marketers (marketer-generated content, i.e., MGC) on consumers' apparel purchase expenditures. A content analysis method was used to construct measures to capture the informative and persuasive nature of UGC and MGC while distinguishing between directed and undirected communication modes in the brand community. In our empirical analysis, we exploit differences across consumers' fan page joining decision and across timing differences in fan page joining dates for our model estimation and identification strategies. Importantly, we also control for potential self-selection biases and relevant factors such as pricing, promotion, social network attributes, consumer demographics, and unobserved heterogeneity. Our findings show that engagement in social media brand communities leads to a positive increase in purchase expenditures. Additional examinations of UGC and MGC impacts show evidence of social media contents affecting consumer purchase behavior through embedded information and persuasion. We also uncover the different roles played by UGC and MGC, which vary by the type of directed or undirected communication modes by consumers and the marketer. Specifically, the elasticities of demand with respect to UGC information richness are 0.006 (directed communication) and 3.140 (undirected communication), whereas those for MGC information richness are insignificant. Moreover, the UGC valence elasticity of demand is 0.180 (undirected communication), whereas that for MGC valence is 0.004 (directed communication). Overall, UGC exhibits a stronger impact than MGC on consumer purchase behavior. Our findings provide various implications for academic research and practice.

Key words: social media; brand community; consumer behavior; user-generated content; marketer-generated content; communication mode; text mining; econometric modeling

History: David Godes, Senior Editor; Wendy Moe, Associate Editor. This paper was received on January 15, 2012, and was with the authors 4 months for 2 revisions. Published online in *Articles in Advance* January 14, 2013.

1. Introduction

Social media have become incredibly popular in recent years. eMarketer projects that more than half of U.S. adult Internet users will be regular users of social media by 2013 (Grau 2009). The number of active Facebook users has already reached 955 million by July 2012, an increase of 29% over the prior year (Facebook 2012). This surge in popularity has produced extensive online user-generated content (UGC) or word of mouth (WOM) and hence, attracted marketers' attention. For instance, more than 1.5 million businesses have set up brand communities (i.e., fan pages) on Facebook for marketing purposes

(Website-Monitoring 2010). Marketers, on behalf of their firms, generate content on social media (hereafter termed as marketer-generated content (MGC)) to engage consumers actively. Despite the prevalent use of social media by consumers and marketers, empirical research investigating their economic values still lags in three critical aspects that motivate our study.

First, prior UGC studies that have documented the economic impact of various aspects of UGC, such as review volume (Chevalier and Mayzlin 2006, Duan et al. 2008, Liu 2006), review subjectivity, and readability (Ghose and Ipeiroitis 2011), have focused

mainly on one-time purchase items or products such as movies (Duan et al. 2008, Liu 2006) and books (Chevalier and Mayzlin 2006). Studies such as Luca (2011) that examine UGC in relation to repeat purchase items are rare, and none have examined both UGC and MGC in the context of a social media brand community. Thus, the literature lacks a rigorous quantification of the value of recurring engagement by consumers and marketers in such a community, especially with metrics such as UGC and MGC elasticities of demand for repeat purchase goods.

Second, prior research has shed little light on the contention between the two complicated roles of consumers and marketers. Even though some research (Chen and Xie 2008, Mayzlin 2006, Trusov et al. 2009) has attempted to evaluate the role of UGC side by side with that of MGC or other marketer actions, empirical evidence on the relative efficacy of UGC and MGC in inducing consumer purchases is rare, with the exceptions of Trusov et al. (2009) and Albuquerque et al. (2012). Because of the simultaneous engagement of consumers and marketers on social media, consumers' purchase decisions are often influenced by both UGC and MGC. The potential conflict stems from different consumer motivations, needs, and at times, their level of skepticism toward MGC (Escalas 2007, Obermiller and Spangenberg 1998). Coupled with the potential two-sidedness (i.e., general positivity and negativity) of interactions from UGC and online WOM (Godes and Mayzlin 2009), it is thus not clear yet in the literature as to what the relative marketing effectiveness of MGC (which typically is overtly positive) and UGC on consumer purchases is.

Third, prior UGC research mostly focused on the aggregate-level economic values of UGC, but overlooked the critical phenomena occurring at the dyadic individual consumer level. Despite the increasing reliance of firms on consumers' WOM as a marketing strategy (Godes and Mayzlin 2009, Nam et al. 2010), little effort has been devoted to understanding whether and how modes of interpersonal communication matter. Consumer-to-consumer communication tends to be *undirected* in the past (e.g., in online reviews), and so does marketer-to-consumer communication propagated in a broadcast manner. Such undirected communications typically address the entire audience base at large without targeting a specific party and without regard for past interactions contexts. However, in social media contexts (e.g., Facebook fan pages), juxtaposed among the undirected communication are often *directed* consumer-to-consumer and marketer-to-consumer communication (Burke et al. 2011). For example, consumers and marketers can pinpoint each other's remarks and respond in a targeted way to each party's content.

They can interact on fan pages on a one-to-one basis via posting or commenting in response to a post. Despite its prevalence, research distinguishing the effects of directed and undirected communication modes of consumers and marketers in affecting consumer behavior still lags.

The objective of our study is to assess the impacts of both UGC and MGC in a social media brand community on consumers' repeat purchase behaviors. By measuring the informative and persuasive aspects of UGC and MGC, and observing them at the dyadic individual consumer level, we seek to quantify their direct and relative impacts under directed and undirected communication modes. Our research question is thus: How is consumer purchase behavior influenced by user-generated content and marketer-generated content in social media brand communities, and whether and how do the communication modes matter?

To answer our research question, we collected UGC and MGC data from an apparel retailer's brand community (i.e., fan page) on Facebook, and matched these with community members' purchase information from the retailer's customer reward program database. We used a commercial text mining tool to construct measures to capture the informative and persuasive nature of UGC and MGC while distinguishing between directed and undirected communication modes in the brand community. Our econometric specification models consumers' weekly purchase expenditure as a function of UGC and MGC factors, controlling for relevant factors at the pricing, promotion, individual consumer, social network, and time unit levels. Our identification strategy for the impacts of UGC and MGC is first based on the propensity score matching technique that enables us to control for self-selection at the fan page level (Moe and Schweidel 2012) via constructing a "control" group of matched customers who were in the reward program but did not join the social media brand community. With the matched customer data sample, we then used a difference-in-differences approach to estimate the economic impact (i.e., "treatment" effect) of joining the brand community. We finally estimated a Heckman selection model to quantify the differential effects of directed and undirected UGC and MGC, while controlling for potential self-selection based on unobserved factors, as well as observed ones such as content generation and network ties. Lastly, we performed robustness checks to validate the consistency of our findings in the presence of potential serial correlation and across differences in time lags and model specifications.

We find evidence that social media brand community contents affect consumer purchase behavior through the embedded information and persuasion.

Importantly, we determine the positive impact of joining the brand community to be about \$25 per consumer. We uncover the different roles played by UGC and MGC in driving consumer purchases, varying by the type of directed or undirected communication modes by consumers and the marketer. Specifically, consumers influence the purchases of one another through both informative and persuasive communications, and marketers influence it only through persuasive communication. Further, undirected contents are more effective than directed ones for both informative and persuasive consumer-to-consumer communication, whereas directed contents are more effective than undirected ones for persuasive marketer-to-consumer communication. The elasticities of demand with respect to UGC's persuasive effect (undirected) and informative effect (directed) are estimated to be 0.180 and 0.006, respectively, whereas that for MGC's persuasive effect (directed) is 0.004. UGC thus exhibits a more influential role than MGC in driving consumer purchases.

Overall, our study makes the following contributions. First, our study unveils the intricate roles of consumers and marketers on social media, and provides a rigorous quantification of the economic impact of a social media brand community's UGC and MGC on consumers' repeat purchases of an apparel brand. Second, our research serves as the first attempt to measure the direct and relative effectiveness and economic values of consumers' online WOM and marketers' proactive marketing activities on social media at the individual consumer level. Third, our findings document the criticality of communication modes of social media content by showing the differential and even contrasting impacts of social media content under directed and undirected communication modes.

2. Literature Review

The popular advent of social media has witnessed a dramatic increase in online engagement and digitalized WOM communication (Dellarocas 2003). Marketers have also capitalized on the trend and launched brand communities on social media platforms to engage consumers and facilitate and generate WOM "buzz" so as to increase information sharing and ultimately drive sales (Kozinets 2002). This has also triggered researchers to investigate the economic value of social media. Early efforts focused on the various outcomes of consumers' engagement in brand communities. For instance, researchers studied consumers' identification (Algesheimer et al. 2005), participation (Bagozzi and Dholakia 2006), and communication (Adjei et al. 2010) in a brand community. They found that these engagements would positively

affect consumers' community participation behavior and commitment, firm trust, and brand purchase behavior.

Other research efforts focused on the online WOM buzz per se, which is the observed output of consumers' engagement on social media. This WOM buzz is typically defined as UGC. Most extant studies focused on the quantitative aspects (e.g., review volume and rating) of UGC and investigated their impact on some aggregate-level¹ economic outcomes. For instance, researchers studied the impact of user-generated reviews on sales of mostly one-time purchase goods, such as movies (Chintagunta et al. 2010, Duan et al. 2008, Liu 2006), books (Chevalier and Mayzlin 2006), video games (Zhu and Zhang 2010), and more rarely, repeat purchase goods such as beers (Clemons et al. 2006) and beauty products (Moe and Trusov 2011). They generally concluded that the quantitative aspects of online reviews such as review volume and/or rating (valence) positively affect aggregate product sales. Apart from online reviews, some studies also examined other types of UGC. Godes and Mayzlin (2004) studied Usenet newsgroup conversations, Tumarkin and Whitelaw (2001) investigated Internet postings in financial discussion forums, Dhar and Chang (2009) studied blog postings, and Albuquerque et al. (2012) studied user-created magazines in an online platform. Likewise, they also reported that quantitative aspects of UGC (e.g., volume, dispersion) were related to aggregate-level economic outcomes.

However, isolated findings on the quantitative aspects of UGC have gradually waned in conclusiveness as the role of qualitative information (e.g., textual content) escalates to the forefront with its importance in the current social media context. For instance, Forman et al. (2008) found that the disclosure of reviewer identity information and a shared geographical location between reviewers and consumers increased product sales, highlighting the impact of qualitative factors. To examine the qualitative aspects of UGC and their economic impact, researchers often use some qualitative analysis methods (e.g., text mining) or tools to extract embedded information from the textual contents. For instance, Pavlou and Dimoka (2006) extracted "benevolence" and "credibility" information embedded in the feedback text comments of sellers on eBay's online auction marketplace. They found that superior past seller performance revealed by the sellers' feedback text comments created price premiums for reputable

¹ Aggregate-level outcomes refer here to metrics such as total sales volume per day and brand market shares, as opposed to individual customer's behavioral outcomes such as purchase expenditure or quantity in a trip or week.

sellers by engendering buyers' trust in the sellers. Gu et al. (2007) extracted the "quality" of postings in virtual communities and found a trade-off between the quality and quantity of postings. Ghose and Ipeirotis (2011) constructed measures for two text-based attributes (subjectivity and readability) of review contents and concluded that these two factors positively affected sales. Additionally, in the finance discipline, Antweiler and Frank (2004) found that the bullishness (sentiment) of messages posted in Internet stock forums helped predict market volatility. Similarly, Das and Chen (2007) identified investor sentiments from stock market message boards and found a relationship between sentiments and stock values. Ghose et al. (2012) leveraged on UGC captured using data-mining techniques from social media platforms to generate a new ranking system for travel search engines. Sonnier et al. (2011) and Tirunillai and Tellis (2012) further classified online communications into positive, negative, and indifferent sentiment categories, and found asymmetric impacts on firm sales and stock trading outcomes. In essence, this stream of studies reported that qualitative aspects of social media UGC exert an impact on aggregate-level economic outcomes.

Despite these research efforts in studying UGC impact, the invariable focus on aggregate-level economic values has resulted in researchers overlooking UGC interpersonal communication at the dyadic individual consumer level. Specifically, UGC captured in past studies tends to be communication in an undirected manner from consumers to consumers. For instance, online reviews (e.g., Chevalier and Mayzlin 2006, Clemons et al. 2006, Duan et al. 2008, Liu 2006) were posted by consumers who have purchased some products, while other consumers who have not purchased or are interested in the products can only read these reviews. However, no directed messages were exchanged because reviewers were essentially writing the reviews with the general public in mind. This also applies to many other types of UGC in past studies, such as financial forums (Tumarkin and Whitelaw 2001) and e-commerce websites (Pavlou and Dimoka 2006). However, social media platforms have now enabled many features for observable, directed interpersonal communication.

There exist only a few studies that examined the relative effect of UGC versus that of MGC, and thus are related to our study. For instance, Mayzlin (2006) developed an analytical model to examine the credibility of online WOM, which can be a mixture of consumer recommendations and disguised firm promotions. She found that consumer WOM can still be persuasive despite the overt promotional intent by firms in such online settings. Chen and Xie (2008) developed analytical models to argue that a

major function of consumer reviews is to serve as a new element in the marketing communications mix. Although they theorized that a firm's decision to provide consumer reviews can increase its incentive to offer more complete product information, there is no relative comparison on the profit impact of consumer reviews and traditional marketing communications. Trusov et al. (2009) studied the effects of WOM marketing on customer acquisition and growth at an Internet social networking site and compared it with traditional marketing mechanisms. This study only focused on aggregate outcomes such as the number of one-time customer acquisitions and not recurring sales by individual customers. The authors obtained a long-term elasticity for online WOM of 0.53, which is about 20 to 30 times higher than that for traditional marketing. Albuquerque et al. (2012) used data from an online user-generated magazine platform to compare content creator activities (e.g., referrals and WOM efforts) with firm-based actions (e.g., public relations). However, they lacked individual customer-specific visitation and communication data, and did not focus on MGC per se nor study qualitative aspects of UGC. Our research differs from the above studies by quantifying the extent to which different aspects of social media content drive sales of a repeat purchase product, in terms of textual aspects (*information richness* and *valence*), and communication modes (*directed* and *undirected*) of types of contents (UGC and MGC) at the *dyadic individual consumer level*.

3. Research Hypotheses

Consumers typically face product uncertainties prior to purchases, so they often seek information from online contents (e.g., consumer reviews) (Chevalier and Mayzlin 2006). Contents from mass media or social media are evaluative and can serve to persuade consumers (Goh et al. 2011). Thus, we aim to examine two effects (*informative effect* and *persuasive effect*²) of UGC and MGC in social media brand community contexts. We focus on two important textual aspects of UGC and MGC, namely, *content information richness* (to capture the *informative effect*) and *content valence* (to capture the *persuasive effect*). Content information richness refers to the amount of information (e.g., product or brand attributes, usage experiences) embedded in the UGC and MGC. Content

² The informative effect of UGC/MGC draws analogy to the notion of informative advertising in the marketing literature, whereby consumers are provided with factual data on the nature and function of the product or service. Correspondingly, the persuasive effect of UGC/MGC parallels the persuasive advertising concept that assumes that consumers already understand the basic function or nature of the product but have to be convinced of the desirability and/or benefits of the product that sets it apart from rival alternatives in a market.

valence refers to the embedded positive or negative sentiment, evaluation, or attitude toward the product or brand, which can be shown through the use of positive or negative words (e.g., good, bad, terrible).

3.1. Content Information Richness

Consumers often face incomplete product information (Kivetz and Simonson 2000), so they need to make purchase decisions under uncertainties (Narayanan et al. 2007, Nelson 1970). As consumers are typically averse to losses (Kahneman and Tversky 1979), they may seek more product-related information to reduce their uncertainties. When uncertainties are reduced, consumers bear more confidence in making purchase decisions (Schubert and Ginsburg 2000). Hence, *ceteris paribus*, when consumers possess more product-related information, they will be more likely to purchase a product that fits their needs or requirements.

A brand community is specialized, because at its center is a branded product (Muniz and O'Guinn 2001). UGC and MGC generated within the community involve product-related information. For instance, UGC may embed consumers' product usage experiences, which involve information of the product (e.g., product features) and other related information (e.g., shopping experiences). MGC may also embed product and other related information (e.g., warranty conditions, after-sales services). As such, we expect information richness of both UGC and MGC to have a positive impact on consumer purchase behaviors.

The comparative impact of UGC and MGC (in terms of the informative effect) is ambivalent. On the one hand, the information asymmetry problem (i.e., firms have complete product information whereas consumers possess incomplete product information) (Akerlof 1970, Mishra et al. 1998) always plagues a consumer-firm relationship. Hence, consumers are tempted to seek information they need from marketers (or representatives of firms), rather than from other consumers who may lack the desired information. As such, MGC information might be more effective than UGC information in addressing consumers' needs and reducing uncertainties. Moreover, search and processing costs are incurred when consumers seek and process information (Ratchford 1982). Because MGC has a higher likelihood to embed information that fits consumers' needs, it will be less costly for consumers' information seeking and processing. As a result, consumers might put more weight on MGC than UGC. Thus, we expect MGC information richness to be more influential than UGC information richness.

On the other hand, there is another school of competing thoughts. Specifically, information generated

by marketers typically describes product information based on technical specifications and is thus product oriented, whereas consumer-generated information tends to describe a product based on usage conditions from a consumer's perspective and is, in contrast, more likely to be consumer oriented (Bickart and Schindler 2001). In other words, UGC information might be more relevant to consumers than MGC information, and thus has the advantage of helping consumers find products matching their preferences (Chen and Xie 2008). This begets the competing hypothesis that UGC information richness will be more influential than MGC information richness in influencing consumer purchases. Summing both perspectives, we arrive at a set of competing hypotheses:

HYPOTHESIS 1A (H1A, COMPETING). *UGC information richness has a smaller impact than MGC information richness on consumers' purchase behavior.*

HYPOTHESIS 1B (H1B, COMPETING). *UGC information richness has a larger impact than MGC information richness on consumers' purchase behavior.*

3.2. Content Valence

Consumers often love to share and relate their product experiences with members of a brand community, expressing their opinions and sentiments (Algesheimer et al. 2005). If consumers are satisfied with a brand or product, they may exhibit favorable attitudes and sentiments toward it. If they dislike the brand or product, or are marred by the experience, they may exhibit negative attitudes and sentiments. Hence, valence embedded in UGC can be interpreted as their general evaluations of a brand or product (Clemons et al. 2006, Liu 2006). Positive (negative) valence of UGC should drive (impede) consumer purchases (Pavlou and Dimoka 2006).

The impact of MGC valence can be discerned from the literature on persuasive advertising (e.g., Russo and Chaxel 2010, Von der Fehr and Stevik 1998). Persuasive advertising involves messages that highlight the positivity of products to enhance evaluations and to instill a sense of good feeling in consumers to tempt them into purchase (Wu et al. 2009). Similarly, marketers embed their positive statements in MGC to create a favorable product reputation and image to influence sales. Hence, we posit that the impact of MGC valence, similar to that of persuasive advertising, positively influences consumers' purchase behavior.

However, MGC may exhibit a weaker persuasive effect than that of UGC. Specifically, over the years, consumers have developed a general tendency to disbelieve or be skeptical toward marketing messages (Escalas 2007). They feel that marketers would resort to gimmicks and tricks (e.g., exaggerating the product benefits while downplaying the weaknesses) in

order to persuade consumers to purchase. In contrast, other consumers have little reasons for doing so. Moreover, consumers tend to trust UGC in evaluating products because they are more similar to one another in terms of community identities, needs, and preferences for specific brands or products and their information (Arazy et al. 2010, Brown and Reingen 1987, Gilly et al. 1998). Thus, consumers might succumb more to UGC persuasion rather than MGC persuasion. Trusov et al. (2009) documented that the impact of user referrals (persuasion) on member growth at an Internet social networking site is higher than that of traditional marketing communications (e.g., media appearances and promotional events). This corroborates our conjecture that UGC might be stronger than MGC in terms of persuasive effect. In essence, we postulate that social media UGC valence has a larger impact than MGC valence in driving purchases.

HYPOTHESIS 2 (H2). *UGC valence has a larger impact than MGC valence on consumers' purchase behavior.*

3.3. Directed Communication versus Undirected Communication

Consumers are inundated with irrelevant information in online environments nowadays (Tam and Ho 2005). Hence, a directed message, which is communicated to a targeted consumer, is expected to be more effective than an undirected one circulated to the mass population, because directed communication easily captures one's attention and elicits a response (Amaldoss and He 2009). Moreover, compared to undirected communication, consumer-to-consumer directed communication is more likely to evoke norms of reciprocity. Such directed communication in brand communities may be more intimate in the message contents such that WOM product recommendation or feedback can be exchanged in a more personalized manner fitting each other's preferences or needs (Burke et al. 2011). We thus postulate that communicating in a directed manner with UGC would be more effective in driving consumer purchases than doing so in an undirected manner for consumer-to-consumer interactions in social media brand communities.

HYPOTHESIS 3 (H3). *For brand community UGC, the impact of directed communication is more effective than that of undirected communication in influencing consumers' purchase behavior.*

The comparative advantage of directed messaging over undirected messaging for MGC communication is equivocal. On the one hand, when marketers directly communicate to a specific consumer, it is easier to capture one's attention relative to undirected communication addressing the entire customer base

without regard for past interaction contexts or specific targeted consumers. Directed marketing messages designed for and communicated to a specific consumer are often tailored to one's needs, heightening the relevance and fit. This ensures that replies can be customized to generate responses or interactions to culminate in eventual purchases (Manchanda et al. 2008). Indeed, directed communications are often exemplary of great customer service.

On the other hand, if marketers frequently engage in unsolicited directed communication with consumers, consumers' skepticism and annoyance (Obermiller and Spangenberg 1998) might be aggravated. This might result in the termination of such communication links (Goh et al. 2011), or disapproving behaviors, such as product boycotts or even the dissemination of negative WOM (Smith and Cooper-Martin 1997). Conversely, undirected marketing communications by a marketer may have a higher level of reach in message receipt by consumers in the brand community of platforms such as Facebook. Undirected communications often get propagated as "posts" or news streams that appear prominently, for instance, on a fan's or consumer's own Facebook "News Feed" page. In contrast, a marketer's directed messages to specific consumers have a lower level of reach or exposure. As such, undirected marketing communication might be more effective than directed communication. Thus, these two camps of arguments give rise to our competing set of hypotheses.

HYPOTHESIS 4A (H4A, COMPETING). *For brand community MGC, the impact of directed communication is more effective than that of undirected communication in influencing consumers' purchase behavior.*

HYPOTHESIS 4B (H4B, COMPETING). *For brand community MGC, the impact of directed communication is less effective than that of undirected communication in influencing consumers' purchase behavior.*

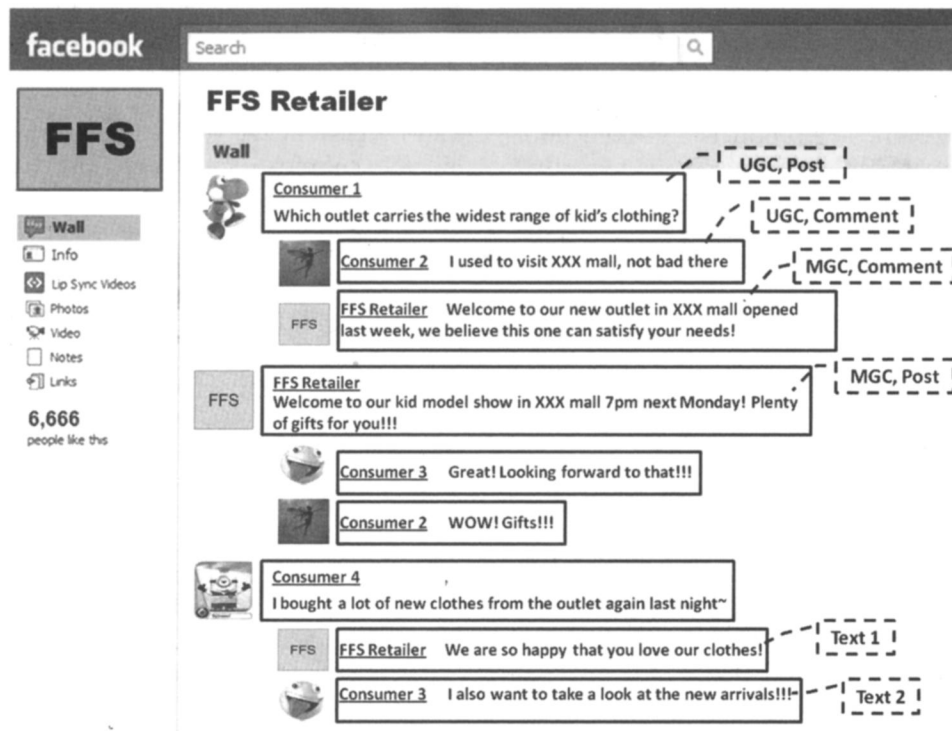
4. Research Methodology

4.1. Research Context

Our research context is a business fan page brand community on Facebook set up in July 2009 by FFS,³ a casual wear apparel retailer in a small Asian market. The retailer also provided us with customer information from their reward program database. Figure 1 presents an edited screenshot of the brand community. FFS retailer set up this community to serve as a platform to engage and interact with their consumers and also to facilitate interactions among consumers.

³ Because of a nondisclosure agreement, we are not able to reveal the identity of the retailer.

Figure 1 FFS Retailer Fan Page Brand Community



Note. The most recent post appears on top, but the most recent comment appears at the bottom of a list of comments related to a particular post.

Consumers can “like” this fan page to engage as community members or fans and then interact with other consumers and the marketer (i.e., FFS retailer). Users interact by generating content, such as posts and comments. Content generated by consumers (or the marketer) are referred to as UGC (or MGC). According to FFS retailer, Facebook is the only social media platform it uses to engage consumers. This thus provides us a thorough, unambiguous setting to examine the impact of UGC and MGC on consumer behavior. Descriptive statistics of the data for this research will be presented in §4.4.

In this community, we observe two types of content, i.e., posts and comments, for both UGC and MGC. Posts are initial text postings that may be addressed to someone (directed) or the entire community (undirected) whereas comments are follow-ups to posts. Although comments are responses to posts, they too can be directed or undirected. Hence, the coders manually read through all posts and comments to ensure the correct coding of communication modes. Posts and comments that were directly addressed to a user are coded as *directed* communications whereas posts and comments that were not directly addressed to a user were deemed as *undirected* communications. For instance, texts 1 and 2 to consumer 4 are directed communications from the marketer and consumer 3, respectively, whereas all other messages generated by

others are considered as undirected communications to consumer 4 (e.g., the phrase “WOW! Gifts!!!” from consumer 2).

4.2. Qualitative Analysis

We employ text mining techniques to analyze the textual or qualitative UGC and MGC data for quantitative analysis. Given a piece of textual content, the text mining tool first decomposes the content into words and phrases based on its large library, and then performs extraction of concepts. Each extracted concept is assigned a corresponding type indicating the sentiment nature (positive, negative, or indifferent).⁴

Because the number of concepts can indicate the richness of information and the type of a concept can reflect the embedded sentiment, our measures of UGC and MGC factors are directly derived from these text mining results. First, information richness is measured as the number of concepts extracted. Previous information extraction studies also extracted information by identifying context-related or context-free concepts (e.g., Rau et al. 1989). Similar approaches have been employed in studies in various disciplines. For instance, researchers had operationalized information

⁴ The online appendix (available at <http://dx.doi.org/10.1287/isre.1120.0469>) elaborates the text mining procedure in detail.

richness as the amount of concepts (e.g., price, quality) communicated by advertisements (e.g., Healey and Kassarian 1983, Resnik and Stern 1977).

Second, valence is measured as the net positivity (i.e., number of positive concepts minus number of negative concepts), which is derived from a sentiment classification algorithm, i.e., *Naïve Classifier* (Das and Chen 2007). Each word in a text is checked against the lexicon and given a value (−1, 0, +1) based on sentiment type (negative, indifferent, positive). The net word count of all lexicon-matched words is taken, and the text is deemed positive (negative) if the value is greater (less) than zero; else, it is indifferent.

4.3. Empirical Model

4.3.1. Communication Intensity. It has been widely acknowledged that online social interactions can allow online users to establish awareness of one another (McKenna et al. 2002), and the awareness may increase with the amount of interactions and eventually lead to online relationship development (Parks and Floyd 1996). Different levels of awareness may result in different levels of communication impact (Brown and Reingen 1987). For instance, one may expect the information from a friend, whom he or she has a higher awareness of, to be more influential compared to the same information from a stranger. In addition, consumers may have a relationship with firms or their representatives such as a marketer, and this relationship may also affect consumers' purchase decisions (Crosby and Stephens 1987). Importantly, trust in online merchants is also typically built up over time with increasing interactions and patronage (Pavlou and Dimoka 2006).

To account for this, we use *communication intensity* to weigh the impact of each *directed* consumer-to-consumer (UGC) and marketer-to-consumer (MGC) communication. Thus, the information richness and valence of each directed communication is weighted by the communication intensity between each pair of communicating users. To account for this intensity between each pair of users, we measure the number of prior directed communications between them, accumulated over time.

4.3.2. UGC Factors. For directed communication, $U_D_IR_{it}$ in Equation (1) and $U_D_VA_{it}$ in Equation (2) denote the average information richness and average valence of UGC that consumer i has observed through directed communications in time period t , where $UDIR_{ijtm}$ and $UDVA_{ijtm}$ are the information richness and valence of the m th UGC that consumer i has observed from consumer j through directed communication in period t . The communication intensity between consumers i and j is denoted as $UIntensity_{ijtm}$, which is measured as the number of previous directed

communications between consumers i and j prior to their m th directed communication in period t . The total number of UGC that consumer j has generated to consumer i through directed messaging in period t is denoted as M_{ijt} . Thus, dividing the inner summation term of weighted $UDIR_{ijtm}$ and $UDVA_{ijtm}$ in Equations (1) and (2) by M_{ijt} obtains the average information richness and average valence of directed UGC from each consumer j . Finally, J_{it} is the total number of consumers who have generated directed messages to consumer i in period t . Therefore, dividing the outer summation term in Equations (1) and (2) by J_{it} derives the mean information richness and valence of directed UGC for consumer i across J_{it} users whom consumer i interacted with in a directed manner:

$$U_D_IR_{it} = \sum_{j=1}^{J_{it}} \left(\frac{\sum_{m=1}^{M_{ijt}} (UDIR_{ijtm} \times UIntensity_{ijtm})}{M_{ijt}} \right) / J_{it} \quad (1)$$

$$U_D_VA_{it} = \sum_{j=1}^{J_{it}} \left(\frac{\sum_{m=1}^{M_{ijt}} (UDVA_{ijtm} \times UIntensity_{ijtm})}{M_{ijt}} \right) / J_{it}. \quad (2)$$

For undirected communication, $U_U_IR_{it}$ in Equation (3) and $U_U_VA_{it}$ in Equation (4) denote the average information richness and valence of UGC that consumer i has observed through undirected communication in period t . The variables $U_U_IR_{it}$ and $U_U_VA_{it}$ are simply the average information richness and average valence of all N_{it} pieces of UGC that consumer i has observed through undirected communication in period t , where $UUIR_{itn}$ and $UUVA_{itn}$ denote the information richness and valence of the n th UGC that consumer i has observed through undirected communication in period t :

$$U_U_IR_{it} = \sum_{n=1}^{N_{it}} UUIR_{itn} / N_{it} \quad (3)$$

$$U_U_VA_{it} = \sum_{n=1}^{N_{it}} UUVA_{itn} / N_{it}. \quad (4)$$

4.3.3. MGC Factors. For directed communication, $M_D_IR_{it}$ in Equation (5) and $M_D_VA_{it}$ in Equation (6) denote the average information richness and average valence of directed MGC that the marketer has communicated to consumer i in period t , where $MDIR_{itr}$ and $MDVA_{itr}$ are the information richness and valence of the r th directed MGC that the marketer has communicated to consumer i in period t . The communication intensity between consumer i and the marketer is denoted as $MIntensity_{itr}$, measured as the number of prior directed communications between consumer i and the marketer prior to their r th directed communication in period t . The

total number of directed MGC that the marketer has communicated to consumer i in period t is denoted as R_{it} :

$$M_D_IR_{it} = \sum_{r=1}^{R_{it}} (MDIR_{itr} \times MIntensity_{itr}) / R_{it} \quad (5)$$

$$M_D_VA_{it} = \sum_{r=1}^{R_{it}} (MDVA_{itr} \times MIntensity_{itr}) / R_{it} \quad (6)$$

For undirected communication, $M_U_IR_{it}$ in Equation (7) and $M_U_VA_{it}$ in Equation (8) denote the average information richness and average valence of MGC that consumer i has observed through undirected communication in period t . The variables $M_U_IR_{it}$ and $M_U_VA_{it}$ are simply the average information richness and average valence of all S_{it} pieces of MGC that consumer i has observed through undirected communication in period t , where $MUIR_{its}$ and $MUVA_{its}$ denote the information richness and valence of the s th MGC that consumer i has observed through undirected communication in period t :

$$M_U_IR_{it} = \sum_{s=1}^{S_{it}} MUIR_{its} / S_{it} \quad (7)$$

$$M_U_VA_{it} = \sum_{s=1}^{S_{it}} MUVA_{its} / S_{it} \quad (8)$$

4.3.4. Control Variables. To obtain robust estimates of the effect of focal UGC and MGC constructs, we control for potentially confounding factors at the pricing, promotion, individual consumer, peer social network, and time levels.

Besides the focal UGC and MGC variables, we also control for other important aspects of UGC and MGC, namely, the volumes of directed UGC ($U_D_VO_{it}$), undirected UGC ($U_U_VO_{it}$), directed MGC ($M_D_VO_{it}$), and undirected MGC ($M_U_VO_{it}$) that consumer i observed in the brand community at period t . To account for potential selection bias at the content generation level, we include variables that measure a user's own posting valence (OWN_VA_{it}) and own posting volume (OWN_VO_{it}), i.e., the average valence and total volume of content generated by consumer i in the brand community at period t .

Importantly, we also include control variables that measure the extent of peer effects, influence and general activity in the FFS brand community, as well as a user's Facebook social network at large. To quantify the influence of a fan, we compute his or her degree centrality⁵ ($CENT_{it}$) on the FFS fan page, based on the communication ties consumer i maintained with other consumers on the fan page in period t . Other

control measures that account for the extent of network ties, activity, and influence from a consumer's Facebook social network at large include the count of Facebook page views⁶ (FB_V_i , i.e., total number of Facebook page views since consumer i 's registration of an account on Facebook), the number of Facebook friends (FB_F_i), and the number of consumer i 's Facebook friends who were also fans on the FFS fan page (FFS_F_i).

To control for the effects of marketing-mix activities, we include a variable $PRICE_t$ that measures the average price (inclusive of discounts) of all products sold in period t . We account for promotional intensity⁷ ($PROM_t$), i.e., the average level of promotion across all days in period t . Promotion on each day is measured as a dummy indicator of a promotional event based on information from the retailer's marketing calendar.

At the consumer level, we account for past expenditure ($PEXP_{it}$), i.e., consumer i 's average expenditure per transaction prior to period t . Other demographic variables captured include a consumer's age⁸ (AGE_i), monthly income (INC_i , i.e., the level of consumer i 's monthly income (1: lowest, 5: highest)), and gender ($MALE_i$, i.e., a dummy indicator for male gender (1: male, 0: female)). Lastly, we include a set of weekly time dummies (θ_t).

4.3.5. Econometric Model Specifications. In Equation (9), we model the influence of UGC and MGC factors on consumers' purchase expenditure. The dependent variable in this study is consumer i 's total purchase expenditure in period t ($EXPEND_{it}$):

$$\begin{aligned} EXPEND_{it} = & \beta_1 U_D_IR_{i,t-1} + \beta_2 U_U_IR_{i,t-1} \\ & + \beta_3 U_D_VA_{i,t-1} + \beta_4 U_U_VA_{i,t-1} \\ & + \beta_5 M_D_IR_{i,t-1} + \beta_6 M_U_IR_{i,t-1} \\ & + \beta_7 M_D_VA_{i,t-1} + \beta_8 M_U_VA_{i,t-1} \\ & + \beta_9 U_D_VO_{i,t-1} + \beta_{10} U_U_VO_{i,t-1} \\ & + \beta_{11} M_D_VO_{i,t-1} + \beta_{12} M_U_VO_{i,t-1} \\ & + \beta_{13} OWN_VA_{i,t-1} + \beta_{14} OWN_VO_{i,t-1} \\ & + \beta_{15} CENT_{i,t-1} + \beta_{16} FB_V_i + \beta_{17} FB_F_i + \beta_{18} FFS_F_i \\ & + \beta_{19} PRICE_t + \beta_{20} PROM_t + \beta_{21} PEXP_{it} \\ & + \beta_{22} AGE_i + \beta_{23} INC_i + \beta_{24} MALE_i \\ & + \theta_t + \alpha_i + \varepsilon_{it}. \end{aligned} \quad (9)$$

⁶ This measure varies across individual consumers but is time invariant, as is the case for FB_F_i and FFS_F_i .

⁷ These marketing promotions were targeted at all consumers (i.e., both fans and nonfans of the community).

⁸ We used consumers' age as of July 2010 (the midpoint of our data set).

⁵ We mapped the network structure of users based on directed content communications on the FFS fan page, i.e., two users or consumers are deemed to be connected to each other if they have ever engaged in directed communications.

We consider UGC and MGC factors in the previous time period ($t - 1$) to avoid simultaneity issues and to allow for a lagged effect from consumers' UGC and MGC exposure to their actual purchases.⁹ β s are the model coefficients of interest, α_i captures unobserved consumer-specific effects, and ε_{it} is the residual error term.

To account for self-selection decisions of consumers joining the FFS brand community, we further specify and estimate a Heckman selection model, i.e., the combination of expenditure model in Equation (9) and selection model in Equations (10) to (12). To model the first-stage fan page selection decision ($BrandCom_i$), we include several exogenous variables as covariates in the first-stage probit model shown in Equations (10) to (12): (1) AGE_i , (2) INC_i , (3) $MALE_i$, two binary indicators of whether a consumer disclosed his or her (4) home phone number ($PHONE_DIS_i$) and (5) home address ($ADDRESS_DIS_i$), and two indicators of whether a consumer opted in to receive promotional information through (6) mobile phone ($PHONE_OPT_i$) and (7) postal mail ($MAIL_OPT_i$) when one signed up as a reward program member.

Selection equation:

$$\begin{aligned} BrandCom_i^* = & \delta_1 AGE_i + \delta_2 INC_i + \delta_3 MALE_i \\ & + \delta_4 PHONE_DIS_i + \delta_5 ADDRESS_DIS_i \\ & + \delta_6 PHONE_OPT_i \\ & + \delta_7 MAIL_OPT_i + \mu_i \end{aligned} \quad (10)$$

$$\begin{aligned} BrandCom_i = 1 & \quad \text{if } BrandCom_i^* > 0, \quad \text{and} \\ BrandCom_i = 0 & \quad \text{otherwise} \end{aligned} \quad (11)$$

$$\begin{aligned} \text{Prob}(BrandCom_i = 1 | z_i) &= \Phi(z_i \delta), \\ \text{Prob}(BrandCom_i = 0 | z_i) &= 1 - \Phi(z_i \delta), \end{aligned} \quad (12)$$

where z_i is a vector of Heckman first-stage model covariates as described in the prior paragraph.

We expect that a consumer's fan page selection decision, $BrandCom_i$, to be related to age, income level, and gender (Muniz and O'Guinn 2001) because FFS is an apparel retailer with trendy, stylish men, women and baby/kids wear offerings. We also expect a user's decision to join the FFS fan page (and thus Facebook) to be related to concerns over data or information privacy (which can be proxied by phone number and address disclosures) and interests in receiving marketing communications from FFS over different channels (Tsai et al. 2011).

4.4. Data Description

The data in our study were drawn from three sources. First, we wrote Java codes based on the Facebook application programming interface to retrieve all user interaction contents from FFS retailer's fan page community on Facebook. Second, Facebook user details and usage logs were obtained from a source related to the Facebook Data Science Team. Third, FFS retailer provided us with (1) the customer reward program database with information for 14,388 customers, (2) the purchase transactions data of customers in this database, and (3) the marketing calendar that detailed the marketing events in a period. These data sets allowed us to construct our major variables of interest and the various control variables. We finally matched Facebook interaction contents data with transactions data by consumer names and organized our model estimation data at the consumer-week level.

Our data spans 104 weeks from when the brand community was first launched in July 2009 till June 2011. By June 2011, the FFS fan page acquired about 6,600 fans in total.¹⁰ On average at the weekly basis, there were about 2.07 MGC posts (std. dev. = 2.08, max = 10) and about 2.59 MGC comments (std. dev. = 3.67, max = 25). Similarly, in terms of UGC participation, the mean UGC postings averaged about 1.62 per week (std. dev. = 2.72, max = 17) and the mean UGC comments averaged around 5.72 per week (std. dev. = 10.11, max = 62). On aggregate, UGC plus MGC participations averaged 12 incidences (std. dev. = 15.57, max = 78) on a weekly basis. In general, we note that there is a high level of heterogeneity or variation in the UGC and MGC contributions on a week to week basis, which provides a vital source of identification for the UGC and MGC effects that can influence purchase behaviors. In assembling the final sample at the consumer-week level, there is no left censoring because we know the date of each fan's joining of the fan page and the date of first purchase.

Our final data sample for model estimations has 398 unique consumers who are both members of the FFS reward program and fans of FFS on the Facebook fan page. Across all purchase transactions, these 398 customers spent on average \$37.05 (std. dev. = \$29.15). We further find that the average purchase expenditure before joining the fan page was \$28.57 (std. dev. = \$29.19), and that after joining the fan page was \$40.52 (std. dev. = \$28.41)—a positive difference of about \$12. Comparatively, the average purchase expenditure for all 14,388 customers in the reward program was \$32.93 across all transactions.

⁹ We compared a set of lag time-period models and determined our choice of one-period lag ($t - 1$) as the best lag level in terms of model fit statistics. The comparison is shown in the online appendix.

¹⁰ To put the number of fans in perspective, we note that the FFS fan page is within the top 100 country-specific Facebook fan pages in terms of acquired fans, as listed on <http://www.socialbakers.com/facebook-pages>.

Table 1 Descriptive Statistics

Variable	Mean	Std. dev.	Min	Max	Median	Skewness
<i>EXPEND</i> (purchase expenditure)	4.711	22.546	0.000	538.420	0.000	8.668
<i>U_D_IR</i> (UGC, directed, information richness)	0.006	0.177	0.000	12.000	0.000	43.621
<i>U_U_IR</i> (UGC, undirected, information richness)	3.143	2.021	0.000	14.000	2.800	0.643
<i>U_D_VA</i> (UGC, directed, valence)	−0.00005	0.019	−1.000	1.000	0.000	−7.706
<i>U_U_VA</i> (UGC, undirected, valence)	0.181	0.539	−3.000	2.000	0.170	0.643
<i>M_D_IR</i> (MGC, directed, information richness)	0.037	0.896	0.000	48.000	0.000	29.184
<i>M_U_IR</i> (MGC, undirected, information richness)	7.010	3.359	0.000	16.000	6.647	0.025
<i>M_D_VA</i> (MGC, directed, valence)	0.004	0.166	−4.000	9.000	0.000	38.177
<i>M_U_VA</i> (MGC, undirected, valence)	0.705	0.987	−2.000	4.000	0.600	0.838
<i>U_D_VO</i> (UGC, directed, volume)	0.026	0.815	0.000	45.000	0.000	40.431
<i>U_U_VO</i> (UGC, undirected, volume)	51.378	172.546	0.000	1,184.000	6.000	5.469
<i>M_D_VO</i> (MGC, directed, volume)	0.004	0.104	0.000	7.000	0.000	36.645
<i>M_U_VO</i> (MGC, undirected, volume)	12.331	21.091	0.000	112.000	5.000	3.036
<i>OWN_VA</i> (Own posting valence)	0.0001	0.013	−0.500	1.000	0.000	65.432
<i>OWN_VO</i> (Own posting volume)	0.003	0.071	0.000	4.000	0.000	27.159
<i>CENT</i> (Degree centrality)	0.0001	0.010	0.000	1.000	0.000	101.002
<i>FB_V</i> (number of Facebook page views)	120.087	148.361	0.000	1,261.000	74.000	3.220
<i>FB_F</i> (number of Facebook friends)	354.254	388.599	0.000	4,791.000	273.000	6.559
<i>FFS_F</i> (number of Facebook friends on FFS)	4.813	6.909	0.000	68.000	3.000	4.811
<i>PRICE</i> (product price)	55.463	18.517	31.036	144.060	50.471	2.355
<i>PROM</i> (promotion intensity)	0.753	0.351	0.000	1.000	1.000	0.351
<i>PEXP</i> (past expenditure)	40.685	28.190	0.000	266.290	38.76	2.200
<i>AGE</i> (age)	32.508	6.216	16.333	54.167	33.229	−0.108
<i>INC</i> (income level)	2.357	0.836	1.000	5.000	2.000	0.884
<i>MALE</i> (gender)	0.110	0.312	0.000	1.000	0.000	2.500

Notes. Observations = 20,406. Mean *EXPEND* across nonzero expenditure weeks = 56.685.

Table 1 shows the descriptive statistics of model variables for the unbalanced panel of 398 consumers across 20,406 observations. A correlation matrix is shown in the online appendix. From Table 1, there is a high level of variability in the UGC and MGC information richness and valence variables, with many cases of overdispersion (i.e., mean > std. dev.). Comparing UGC with MGC, the means and standard deviations of MGC information richness and valence variables are higher than those of equivalent UGC variables.¹¹

5. Model Estimation and Results

5.1. Identification Strategies

Our first identification strategy for the impacts of UGC and MGC is based on the propensity score matching (PSM) method¹² (Heckman et al. 1998,

Rosenbaum and Rubin 1983). This enables us to control for self-selection at the fan page level (Moe and Schweidel 2012) via constructing a “control” group of matched 398 customers¹³ who were in the reward program but did not join the FFS brand community. The major difference between these two groups is that consumers in the “treatment” group were fans on FFS retailer’s Facebook fan page and thus could get exposed to UGC and MGC, whereas those in the control group were not fans and thus had no exposure to UGC or MGC. Given that consumers across the control and treatment groups were essentially identical to one another across the set of exogenous variables (age, income, gender, home phone and address disclosures, mobile phone, and mail opt-ins for marketing information) used as the criteria for matching, self-selection at the fan page level based on these observed attributes is thus controlled for (see online appendix for details). These sets of consumer attributes are comprehensive and informative, such that they influence the treatment assignment (i.e., joining the fan page) and yet are not affected by the treatment, thus satisfying the unconfoundedness or selection on observables identification assumption

¹¹ Although it may appear counterintuitive that there are high variability and negative values in MGC valence, it can be explained by instances where some consumers requested for home delivery services, but the marketer had to apologize for the unavailability of such services. Some consumers also complained about poor in-store services, and the marketer apologized while offering discount coupons as compensation. Such compensatory marketer actions may overreact at times in order to maintain customer satisfaction levels, thus explaining the higher means and variability of MGC factors.

¹² The online appendix elaborates the matching procedure in detail, and explains how our rich data context satisfies the identification

assumptions of unconfoundedness (or selection on observables) needed for this matching method.

¹³ The average purchase expenditure for these 398 propensity-score matched customers was \$34.63.

of PSM. PSM however does not allow for selection on unobservables (which our next two identification strategies allow), and thus can only match based on observed attributes, but not unobserved, potentially confounding factors.¹⁴ Another limitation is that PSM can only estimate treatment effects where there is support for the treated individuals among the “non-treated” population. Lastly, as is the case with other partial equilibrium evaluation methods, PSM cannot establish the impact of the treatment beyond the eligible group of consumers.

With the matched customer data sample, our second identification strategy exploits differences across consumers’ fan page joining decision and across timing differences in fan page joining dates to use a difference-in-differences (DID) model estimation approach. This thus enables us to estimate the economic impact (i.e., treatment effect) of joining the FFS brand community. Although our data context construes an appropriate identification strategy using the DID approach that allows for selection on (time-invariant) unobservables, there are limitations to this method. First, the DID approach is valid only when the treatment is as good as random when conditioned on individual, group, and time fixed effects. Second, the validity of DID estimates may be threatened by the potential endogeneity of the treatments or interventions themselves (e.g., in our context, if loyal consumers have a time-varying propensity to join the retailer’s fan page). Lastly, DID model estimations may be susceptible to serial correlation problems (Bertrand et al. 2004).

Furthermore, with the same matched data sample, our third identification strategy uses a Heckman selection model to quantify the effects of directed and undirected UGC and MGC, while controlling for potential self-selection at other levels such as content generation and network ties, or that associated with unobserved factors. The Heckman selection model takes on specific normal distribution assumptions for the unobservable characteristics that jointly influence the fan page selection decision and the purchase outcome. The estimated model parameters may thus be sensitive to these distributional assumptions of the residuals that provide a technical basis of the Heckman model’s identification (which need not rely strictly on the variation in the explanatory variables).

¹⁴ Sensitivity tests to check on potential deviations from unconfoundedness reveal that the Γ cutoff value is 1.3 (1.35) before an upper bound of significance value reaches above 0.05 (0.10). This implies that to attribute a higher level of purchase expenditure because of an unobserved covariate, rather than to joining FFS’s fan page, that unobserved covariate would need to produce a 30%–35% increase in the odds of joining FFS’s fan page. This thus quantifies the extent of insensitivity of our PSM results to biases from potential unobserved factors.

Another limitation is that model estimation results are unreliable if there are no exclusion restrictions (i.e., at least one exogenous independent variable from the first-stage selection model is excluded from the set of independent variables for the second-stage model).

5.2. Preliminary Analysis and Results

Prior to estimating our main model specification shown in Equation (9), we first conduct a preliminary analysis using a baseline alternative model with a series of main effects and interactions between the four variables of the source of content (UGC/MGC), directed/undirected communication, content information richness, and valence. This preliminary analysis seeks to examine the impact of information richness (IR) and valence (VA) of social media brand community contents on consumer purchase behavior, and then further investigates how IR and VA depend on content source (SOURCE, i.e., UGC volume/MGC volume ratio) and communication mode (MODE, i.e., directed content volume/undirected content volume ratio).¹⁵

We first estimate a model with only the four main effect variables (plus other control variables), using both a fixed effects (FE) and a random effects (RE) specification. The main effects model estimation results reveal significant positive main effects of IR and VA that are consistent with prior studies on online WOM. Next, we follow up with estimating a model with both the main effects and interaction effects variables, and find a significant main effect of VA and also importantly, a significant interaction effect of SOURCE * MODE (see the online appendix for detailed model estimation results). This significant interaction coefficient thus indicates the importance of content source and communication mode, providing support to investigating content source and communication mode in brand communities according to the main model specification given in Equation (9).

5.3. Main Analysis and Results

In our main analysis, we first estimate a FE model and a RE model of consumers’ purchase expenditure (EXPEND) on all control variables that have been

¹⁵ The baseline alternative model specification in the preliminary analysis is

$$\begin{aligned} \text{EXPEND}_{it} = & \beta_1 \text{IR}_{i,t-1} * \text{SOURCE}_{i,t-1} * \text{MODE}_{i,t-1} \\ & + \beta_2 \text{VA}_{i,t-1} * \text{SOURCE}_{i,t-1} * \text{MODE}_{i,t-1} \\ & + \beta_3 \text{IR}_{i,t-1} * \text{SOURCE}_{i,t-1} + \beta_4 \text{IR}_{i,t-1} * \text{MODE}_{i,t-1} \\ & + \beta_5 \text{VA}_{i,t-1} * \text{SOURCE}_{i,t-1} + \beta_6 \text{VA}_{i,t-1} * \text{MODE}_{i,t-1} \\ & + \beta_7 \text{SOURCE}_{i,t-1} * \text{MODE}_{i,t-1} + \beta_8 \text{IR}_{i,t-1} + \beta_9 \text{VA}_{i,t-1} \\ & + \beta_{10} \text{SOURCE}_{i,t-1} + \beta_{11} \text{MODE}_{i,t-1} + \text{ControlVariables} \\ & + \alpha_i + \varepsilon_{it}. \end{aligned}$$

widely recognized as important factors affecting consumer purchase behavior. As reported in Table 2, columns (1) and (2), a few control variables such as prior purchase expenditure and UGC volumes have explanatory power.¹⁶

Next, before we examine the impact of the various UGC and MGC factors of interest, we estimate a DID model to compare consumer purchase expenditure between fans and nonfans, as well as before and after becoming a fan of FFS brand community. Specifically, we created an estimation data sample of 796 consumers, combining the 398 PSM-matched consumers with the original 398 consumers who were fans of the FFS fan page. We use a binary variable, *BrandCom*, to indicate whether each of the 796 consumers was a fan in the brand community (1: fan, 0: nonfan). We then use an additional binary variable, *BecomeFan*, to indicate the timing of becoming a fan (1: after, 0: before) for the 398 fans, and interact it with *BrandCom* (i.e., *BrandCom * BecomeFan*). Because *BrandCom* and *BecomeFan* might be endogenous, we first use several exogenous variables (*AGE*, *INC*, *MALE*, *PHONE_DIS*, *ADDRESS_DIS*, *PHONE_OPT* and *MAIL_OPT*) in a probit model to model the outcome of an unobserved latent variable determining the selection decisions. We thus estimate a treatment effects (TE) model focusing on the coefficient for *BrandCom * BecomeFan*, while controlling for the various control variables. As shown in Table 2, column (3), the DID parameter estimate is 24.597 (± 2.040), which is significantly positive. This implies a significant positive impact of about \$24.60 in purchase expenditure after joining the brand community of FFS retailer. The exposure to UGC and MGC thus has a significant impact on purchase behavior, which gives credence to further explore the impact of different UGC and MGC factors in depth.

We further estimate a full FE model, including all the UGC and MGC factors of focal interest. Table 2, column (4), reports the results. For UGC factors, both information richness and valence are found to have a significant impact on *EXPEND*. Specifically, the coefficients of *U_D_IR* (3.225 ± 1.863), *U_U_IR* (21.849 ± 7.994), and *U_U_VA* (76.733 ± 33.224) are positive and statistically significant. For MGC factors, only valence, i.e., *M_D_VA* (3.383 ± 1.607) is found to have a positive and significant impact on *EXPEND*. Next, we further estimate a full RE model. In Table 2, column (5), the RE model shows similar results to those in column (4). The Hausman test suggests that the

RE estimates are not inconsistent ($\chi^2 = 0.69$, $p = 0.99$). Nevertheless, we prefer the FE model over the RE one because the former allows the consumer-specific unobserved heterogeneity to be correlated to the observed variables (i.e., a more tenable assumption), and its estimation involves a conditional analysis restricted to a specific sample (thus matching our data from the FFS reward program).

Both the prior FE and RE model estimation results have not accounted for potential self-selection at the fan page level. To control for self-selection as a potential confounding factor in determining the effects of consumers' exposure to UGC and MGC on their purchase behavior, we use as model estimation sample, the PSM-matched 398 nonfan consumers as a control group in addition to the original 398 fans. We use *BrandCom* to indicate whether each of the 796 consumers was a fan in FFS retailer's fan page brand community. We then employ the Heckman two-step selection model (Heckman 1976, 1979), including a full set of exogenous consumer-specific covariates in the first step to model the selection decision.¹⁷ In the second step, besides the focal UGC and MGC factors and control variables, we also include consumer fixed effects to account for consumer heterogeneity in the purchase expenditures. As indicated in Table 2, column (6), the estimates are consistent with those in the FE model. Specifically, the parameter estimates for the focal UGC and MGC factors of *U_D_IR* (3.182 ± 1.838), *U_U_IR* (21.317 ± 7.891), *U_U_VA* (74.311 ± 32.819), and *M_D_VA* (3.372 ± 1.570) are all statistically significant. Thus, the information richness of both directed and undirected UGC have a positive influence on consumer purchase expenditure, but not for the case of MGC. In terms of content valence, the valence of directed MGC has a positive effect on expenditure whereas that for directed UGC does not. However, whereas the valence of undirected UGC has a large positive effect on expenditure, there is no effect of undirected MGC at all.

Importantly, the mean of all the four statistically significant UGC and MGC parameter estimates average to about 25.546, which is very close to the DID *BrandCom * BecomeFan* parameter estimate of 24.597 from Table 2, column (3). To further establish the robustness of results from the Heckman selection model on the PSM-matched data sample, we also estimate the Heckman model using consumers from the rest of the entire customer reward program database (i.e., the 13,990 nonfan consumers) as the control group.¹⁸ We report in Table 2, column (7) results that are consistent with those in column (6).

¹⁶ The model fit statistics R^2 of the fixed and random effects models shown in Table 2 are generally small. This is because our estimation data have many zero-expenditure weeks of each customer. Dropping these zero-expenditure weeks increases the R^2 of the estimated models to about 0.107 to 0.139, but this would omit relevant UGC and MGC information that may bias the results. Our research does not involve forecasting, thus R^2 model fit may matter less.

¹⁷ Results of the first-step probit model estimation are shown in the online appendix.

¹⁸ We cannot include all consumer fixed effects because of PC memory limitations with the large number of consumers.

Table 2 Model Estimation Results

Variable	(1) FE Control	(2) RE Control	(3) DID PSM, TE	(4) FE Full	(5) RE Full	(6) Heckman PSM, FE	(7) Heckman Population
<i>U_D_IR</i> (UGC, directed, information)				3.225* (1.863)	3.195* (1.849)	3.182* (1.838)	3.523* (1.873)
<i>U_U_IR</i> (UGC, undirected, information)				21.849*** (7.994)	22.042*** (7.977)	21.317*** (7.891)	22.973*** (8.105)
<i>U_D_VA</i> (UGC, directed, valence)				6.641 (9.009)	6.195 (8.996)	6.603 (8.883)	5.290 (9.138)
<i>U_U_VA</i> (UGC, undirected, valence)				76.733** (33.224)	77.793** (33.151)	74.311** (32.819)	81.355** (33.708)
<i>M_D_IR</i> (MGC, directed, information)				−0.437 (0.389)	−0.422 (0.387)	−0.448 (0.386)	−0.400 (0.393)
<i>M_U_IR</i> (MGC, undirected, information)				−14.209 (22.570)	−13.352 (22.526)	−15.882 (22.493)	−11.962 (23.118)
<i>M_D_VA</i> (MGC, directed, valence)				3.383** (1.607)	3.234** (1.600)	3.372** (1.570)	2.800* (1.606)
<i>M_U_VA</i> (MGC, undirected, valence)				71.473 (86.292)	66.878 (86.095)	76.714 (84.069)	59.933 (86.379)
<i>BrandCom * BecomeFan</i> (DID treatment effect)			24.597*** (2.040)				
<i>U_D_VO</i> (UGC, directed, volume)	0.751* (0.410)	0.798* (0.408)		0.910** (0.462)	0.959** (0.460)	0.917** (0.456)	1.101** (0.468)
<i>U_U_VO</i> (UGC, undirected, volume)	0.199 (0.233)	0.222 (0.232)		0.089 (0.249)	0.114 (0.248)	0.099 (0.245)	0.157 (0.250)
<i>M_D_VO</i> (MGC, directed, volume)	−6.810** (3.331)	−3.655 (2.382)		−8.772 (5.410)	−7.504 (17.361)	−6.771 (17.176)	−8.303 (17.644)
<i>M_U_VO</i> (MGC, undirected, volume)	−2.059 (2.484)	1.294 (0.841)		0.559 (2.954)	1.967 (16.576)	2.371 (16.401)	2.057 (16.855)
<i>OWN_VA</i> (Own posting valence)	−4.845 (12.260)	−4.347 (12.257)		9.672 (13.646)	10.372 (13.639)	9.138 (13.470)	11.900 (13.891)
<i>OWN_VO</i> (Own posting volume)	9.443*** (3.054)	9.528*** (3.048)		4.826 (3.363)	4.927 (3.357)	4.908 (3.313)	5.079 (3.406)
<i>CENT</i> (Degree centrality)	−3.252 (15.716)	−2.716 (15.552)		−10.008 (16.183)	−9.348 (16.021)	−10.009 (15.984)	−10.401 (16.231)
<i>FB_V</i> (No. of Facebook page views)		−0.001 (0.002)			−0.002 (0.002)	0.076 (0.126)	−0.001 (0.001)
<i>FB_F</i> (No. of Facebook friends)		−0.001 (0.001)			−0.001 (0.001)	0.227 (0.275)	−0.001** (0.000)
<i>FFS_F</i> (No. of Facebook friends on FFS)		0.038 (0.047)			0.037 (0.047)	6.402 (7.895)	0.054** (0.024)
<i>PRICE</i> (Product price)	0.188 (0.237)	0.153 (0.136)	0.107 (0.065)	1.171 (1.764)	1.240 (1.555)	1.337 (1.556)	1.214 (1.600)
<i>PROM</i> (Promotion intensity)	5.983 (22.780)	−68.137 (50.279)	−32.551* (18.589)	38.603 (261.978)	−670.859 (877.171)	−703.456 (863.461)	−649.995 (887.737)
<i>PEXP</i> (Past expenditure)	−0.029*** (0.011)	−0.002 (0.008)	0.034*** (0.006)	−0.029*** (0.011)	−0.002 (0.008)	−0.029*** (0.011)	0.028*** (0.006)
<i>AGE</i> (Age)		0.017 (0.051)	0.045 (0.068)		0.017 (0.052)	−5.370 (6.312)	0.033 (0.028)
<i>INC</i> (Income level)		−0.117 (0.386)	0.305 (0.500)		−0.113 (0.387)	310.863 (375.315)	0.022 (0.201)
<i>MALE</i> (Gender)		0.023 (0.963)	−0.981 (1.238)		0.022 (0.966)	−553.704 (673.973)	0.060 (0.513)
<i>Constant</i>	−12.676 (23.447)	35.117 (22.325)	12.051 (11.118)	−200.030 (125.129)	481.263 (638.524)	0.000 (0.000)	455.053 (643.396)
<i>Time dummies</i>	−included- 398	−included- 398	−included- 796	−included- 398	−included- 398	−included- 796	−included- 14,388
Number of consumers	20,406	20,406	61,160	20,406	20,406	52,250	840,708
Hausman test/selection ρ	$\chi^2 = 8.36, p = 0.99$			$\chi^2 = 0.69, p = 0.99$		$\rho = 0.000$	$\rho = -0.066$
R^2	0.0240	0.0273		0.0246	0.0279		
Wald χ^2		600.00	871.46		613.78	3,198.25	612.30

Note. Standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Noteworthy, the Heckman model accounts for selection on unobservables and also potential selection at the content generation and consumption level, because it includes control variables of a fan's own posting valence (*OWN_VA*) and volume (*OWN_VO*) in the brand community, as well as a fan's number of Facebook page views (*FB_V*). Finally, our model attempts to account for selection at the network-tie or peer influence level by including control variables associated with the social network circles of a consumer, i.e., a fan's network degree centrality based on interactions solely on the FFS fan page (*CENT*), number of Facebook friends (*FB_F*), and number of Facebook friends who were also in the FFS brand community (*FFS_F*). Therefore, the above controls give further credence to the impacts of UGC and MGC information richness and valence on purchase behavior, after having accounted for observed and unobserved potentially confounding factors.

In summary, we consider the Heckman two-step selection model based on the PSM-matched control group (Table 2, column (6)) as our best model, because it accounts for selection bias and consumer-specific heterogeneity. To compare the relative impact of UGC and MGC in terms of information richness and valence, and also the relative impact of directed and undirected communication modes, we report the marginal effects and elasticities for the significant UGC and MGC factors in Table 3 based on the main model. We summarize our hypotheses testing results in Table 4. For information richness, only UGC factors, *U_D_IR* (marginal effect = 3.182, $p < 0.1$) and *U_U_IR* (marginal effect = 21.317, $p < 0.01$), are significant, thus supporting H1B and rejecting its competing Hypothesis H1A. For valence, the significant marginal effect of the UGC factor, *U_U_VA* (marginal effect = 74.311, $p < 0.05$), is more than 22 times that of the only significant MGC factor, *M_D_VA* (marginal effect = 3.372, $p < 0.05$), thus supporting H2. Finally, as for directed and undirected communication modes, UGC information richness and valence are generally significant and with larger marginal effects in the undirected mode, thus rejecting H3. On the contrary, for MGC, valence is significant in the directed communication mode only, thus supporting H4A and rejecting its competing Hypothesis H4B.

Table 3 Marginal Effects and Elasticities

UGC factors	<i>U_D_IR</i>	<i>U_U_IR</i>	<i>U_D_VA</i>	<i>U_U_VA</i>
Marginal effect	3.182*	21.317***	—	74.311**
Elasticity	0.006*	3.140***	—	0.180**
MGC factors	<i>M_D_IR</i>	<i>M_U_IR</i>	<i>M_D_VA</i>	<i>M_U_VA</i>
Marginal effect	—	—	3.372**	—
Elasticity	—	—	0.004**	—

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 Hypotheses Testing Results

Hypothesis		Support
H1A, competing	UGC information richness < MGC information richness	No
H1B, competing	UGC information richness > MGC information richness	Yes
H2	UGC valence > MGC valence	Yes
H3	UGC: directed communication > undirected communication	No
H4A, competing	MGC: directed communication > undirected communication	Yes
H4B, competing	MGC: directed communication < undirected communication	No

5.4. Robustness Checks

We further corroborate our main findings by checking its robustness in multiple ways. For ease of reference, Table 5, column (1), presents the main results from Table 2, column (6). For brevity, from this point onward, we only report the major variables of interest for hypotheses testing.

First, we examine the effects of UGC and MGC factors without accounting for communication intensity. We remove all intensity elements from Equations (1), (2), (5), and (6) in §4, and compute only the average UGC and MGC factors for directed communication. Table 5, column (2) shows the model estimates, which qualitatively remain consistent with our main results. However, the comparatively larger and thus potentially misleading coefficient size of the *M_D_VA* parameter (i.e., 7.234), relative to those from all the other models, highlights the importance of accounting for communication intensity.

Second, we check the robustness of our findings across different model specifications. We first estimate a population-averaged (PA) model that allows for an exchangeable correlation structure of a generalized linear model, and then a random effects model estimated via maximum likelihood (RE-ML). The corresponding results for the PA and RE-ML models are shown in Table 5, columns (3) and (4). The model parameter estimates remain consistent with those of the main one in column (1).

Next, to account for the existence of potential serial correlation, we estimate a FE model with a first-order autoregressive disturbance structure (FE-AR1). As indicated in Table 5, column (5), the model estimates under an AR1 structure are consistent with those of the main one. This implies that findings from our main model in column (1) are robust to serial correlation. Lastly, robustness checks on the covariate time lags of the main Heckman model and FE model are detailed in the online appendix. In sum, we are confident of the robustness of our findings given that various checks indicated robustness and consistency.

Table 5 Robustness Checks

Variable	(1) Main	(2) Intensity	(3) PA	(4) RE-ML	(5) FE-AR1
<i>U_D_IR</i> (UGC, directed, information)	3.182* (1.838)	3.382* (1.829)	3.193* (1.846)	3.190* (1.843)	3.531* (2.094)
<i>U_U_IR</i> (UGC, undirected, information)	21.317*** (7.891)	21.658*** (7.904)	22.035*** (7.963)	22.012*** (7.950)	22.748*** (8.104)
<i>U_D_VA</i> (UGC, directed, valence)	6.603 (8.883)	4.938 (8.759)	6.203 (8.981)	6.229 (8.966)	7.804 (9.006)
<i>U_U_VA</i> (UGC, undirected, valence)	74.311** (32.819)	75.959** (32.879)	77.761** (33.094)	77.659** (33.039)	85.498** (33.801)
<i>M_D_IR</i> (MGC, directed, information)	−0.448 (0.386)	−0.861 (0.865)	−0.422 (0.386)	−0.423 (0.386)	−0.467 (0.388)
<i>M_U_IR</i> (MGC, undirected, information)	−15.882 (22.493)	−12.976 (22.407)	−13.369 (22.487)	−13.426 (22.449)	−6.860 (22.920)
<i>M_D_VA</i> (MGC, directed, valence)	3.372** (1.570)	7.234* (3.979)	3.237** (1.598)	3.246** (1.595)	2.764* (1.645)
<i>M_U_VA</i> (MGC, undirected, valence)	76.714 (84.069)	81.518 (89.616)	66.958 (85.947)	67.220 (85.802)	30.434 (89.035)
<i>Constant</i>	0.000 (0.000)	0.000 (0.000)	481.495 (637.421)	482.251 (636.340)	418.512 (608.180)
<i>Control variables</i>	-Included-	-Included-	-Included-	-Included-	-Included-
Number of consumers	796	796	398	398	398
Number of observations	52,250	52,250	20,406	20,406	20,009
Wald χ^2	3,198.25	3,196.86	615.95	608.85	—

Note. Standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6. Discussion and Contribution

6.1. Discussion of Findings

Our study that investigates the impact of social media brand community contents on consumer purchase behavior has several notable findings. First, we empirically show that engagement in social media brand community leads to a significant increase in consumer purchases. Second, our in-depth examination of community contents (UGC and MGC) attests to the fact that brand community contents affect consumer purchase behavior through embedded information as well as persuasion. Besides UGC, MGC in communities also matter, but differently, in influencing consumer purchases. Consumers influence the purchase expenditure of one another through both informative as well as persuasive interactions, whereas marketers influence it only through persuasive communication. Interestingly, consumers' persuasive effect is more than 22 times that of marketer's in terms of marginal effect. The elasticities of demand with respect to UGC's valence (undirected) and information richness (directed) are estimated to be 0.180 and 0.006, respectively, whereas that for (directed) MGC's valence is 0.004. Overall, UGC exhibits a more influential role than MGC in driving purchases.

Finally, evidence affirms directed communication and undirected communication matter differently for UGC and MGC. Specifically, in driving purchases, undirected contents are more effective than

directed ones for both informative and persuasive consumer-to-consumer communication, whereas directed contents are more effective than undirected ones for persuasive marketer-to-consumer communication. For the rejected Hypothesis H3, a plausible reason might be due to the manner that posts and comments on Facebook are structured or displayed. UGC undirected communications typically appear as posts on a fan page with the most recent post appearing in the most salient top-most position that can garner the most attention. In contrast, comments are sorted in the opposite manner with the most recent one listed at the bottom.

6.2. Theoretical Contributions

Our study contributes to the discourse on social media in the following ways. First, the predominant emphasis of prior brand community research on consumer engagement and content (i.e., consumer side; e.g., Algesheimer et al. 2005, Bagozzi and Dholakia 2006, Porter and Donthu 2008) may have unwittingly resulted in the misconception that businesses can only passively react. By accentuating the role of MGC and its impact (i.e., marketer side), we underscore that marketers can actually transform their role from a passive and reactive one to a proactive and influential one. By actively engaging consumers in brand communities, marketers can better reap economic values from social media brand communities.

Second, by juxtaposing the role of MGC besides that of UGC, we unravel the contention and intricacies between the two, thereby complementing and enriching past works. Our findings suggest that MGC does affect consumer purchase behavior, but in a different way from UGC (e.g., Dhar and Chang 2009, Sonnier et al. 2011, Tumarkin and Whitelaw 2001). Hence, the sole reliance on UGC to explain consumer behavior would overlook and omit the persuasive effect of the marketer's social media contents. The differential and even contrasting impact of UGC and MGC suggests that consumers not only respond to the information of online contents, but also factor the sources of content into consideration. This provides a foray into better understanding the economic value of content on social media platforms.

Third, as one of the pioneer efforts to quantify the economic impact of both UGC (or online WOM) and MGC (marketers' proactive marketing activities) on social media, we augment the discourse on social media marketing with insights on its return on investment (ROI). Using various identification strategies, we provide a rigorous estimate of the consumer's economic impact of joining social media brand communities. Our attempt is also one of the first to empirically quantify the relative effectiveness of UGC and MGC in social media brand community contexts.

Fourth, our research is also among the first to propose and validate a model to quantify the economic impact of social media brand community contents at the individual consumer level. This approach enables us to control for consumer heterogeneity selection biases and to address the prior overlooked impact of dyadic communication in terms of the communication modes. Our findings underscore that sharing information alone in brand communities is a necessary but not sufficient condition to generating positive economic outcomes. In addition to contents per se, whether contents are communicated in a directed or undirected manner matters.

6.3. Practical Implications

Our study has several important practical implications to social media marketers. Consumers (UGC) play both informative and persuasive roles and marketers (MGC) play a persuasive role in social media contexts. This suggests that a mere reliance on marketers' own marketing activities may not be the most effective way to drive consumer purchases. Similarly, marketers' total reliance on consumers' WOM buzz is also suboptimal. An ideal strategy would be the right combination of both UGC and MGC. Apart from marketers' diligent preparation of their own persuasive content (e.g., use more favorable or positive words and phrases to describe products and services), marketers should conscientiously

design campaigns to encourage informative, and especially, persuasive communication among consumer themselves on social media platforms. For instance, marketers can incentivize consumers to share their experiences by using discount coupons and reward points. Not unheard of, there are also marketers who employ a community manipulation strategy (Dellarocas 2006) by anonymously behaving as "fellow consumers" to share positive product information within communities.

Second, in social media contexts, directed messaging is more effective for persuasive marketer-to-consumer communication, whereas undirected messaging is more effective for informative and persuasive consumer-to-consumer communication, in driving consumer purchases. Hence, when communicating persuasive content to consumers, marketers can choose a directed communication mode for higher ROI impact. In other words, they can generate content to a targeted user or group for better consumer responses. For instance; in the context of Facebook, a marketer can direct marketing communication in the "comment" entries of the fan page to address specific consumers. With regard to informative and persuasive communications among consumer themselves, marketers can encourage consumers to engage more in undirected communications. For instance, marketers can reward consumers who are most active in sharing their content in posts addressed to the fan page members at large.

Third, marketers might want to enhance their analytics by moving beyond the traditional insights from quantitative analysis, such as the identification of the advertising expenditure-sales relationship, to embrace more insights from qualitative analysis as well. Currently available qualitative tools such as the one we adopted can help track, analyze, and enlighten the content embedded within UGC in their brand communities. Marketers can then get a more nuanced understanding of consumers' general response, attitude toward, and evaluation of the products and marketing campaigns launched.

Finally, our study also presents implications for the design of social media marketing platforms. Many current platforms (e.g., Yelp.com) are popular, and have attracted extensive information sharing in the form of reviews from consumers. However, these platforms do not currently provide much access to marketers' proactive engagements. Indeed, our study suggests that these platforms can actually do better by enabling marketers' engagements. For instance, apart from displaying consumer reviews of a restaurant, social media platforms can also provide free or paid access to marketers from a restaurant to communicate marketing information (e.g., introduction of new cuisines, replies to customers' queries) and to

integrate functional aspects of customer relationship management within the social media platform.

7. Conclusion

Although this research has highlighted several notable findings, we acknowledge some limitations. First, our research context does not entail randomized trials or field experimentations on the UGC and MGC constructs of interest. As such, although we spent considerable efforts in addressing concerns related to selection biases (because of both observables and unobservables), our identification strategies centering on PSM and the Heckman selection model only afford us a quasi-counterfactual of a consumer being a brand's fan on Facebook, after accounting for selection on observables and unobservables.¹⁹ We have also discussed at length previously in §5.1 the limitations of each identification strategy we used. Second, apart from textual contents, there were a small number of pictures and videos in our research context. These contents were posted together with some textual descriptions, which at the same time were captured in our sample. Although we were able to account for the impact from all textual contents, we did not account for the other types of content. Third, the data sample for our research context comes from only a single retailer and its consumers as well as brand community members. Nevertheless, the phenomenon of UGC and MGC interactions is not unique to the FFS community on Facebook.²⁰ Moreover, in terms of the platform used, many other social media platforms (e.g., MySpace, YouTube) offer similar functionalities for marketers and consumers to engage in social interactions. Reassuringly, the parent retail company of the FFS retailer is well established as a franchisee of many famous global apparel brands and thus follows both industry recommended practices and brand-guided procedures with regard to social media marketing communications.

Moving forward, we present potential avenues for future research. A meaningful extension to this research is to investigate the role of product type, perhaps in a randomized trial or experimentation setting (Aral and Walker 2011). As discussed, UGC is more consumer oriented relative to MGC. This may potentially contribute to the stronger role of UGC relative to MGC in our context where experience products

(i.e., apparels) were studied. To what extent do our findings apply to search products (e.g., books, plane tickets) context deserves further scrutiny. It might also be worthwhile to study the relative effectiveness of online (UGC and MGC) and offline marketing activities concurrently. Because firms often face limited marketing resources in multichannel marketing settings (Chu et al. 2007, Zhang 2009), assessing their relative effectiveness and identifying the optimal combination of marketing strategies across multiple channels to achieve better sales outcome is vital.

Electronic Companion

An electronic companion to this paper is available as part of the online version at <http://dx.doi.org/10.1287/isre.1120.0469>.

Acknowledgments

The authors thank the senior editor, associate editor, the anonymous reviewers, Ivan Png, Ke-Wei Huang, Yuanyuan Chen, Seung Hyun Kim, Tuan Quang Phan, the seminar and conference participants at the *Information Systems Research Special Issue Workshop* at the University of Maryland, and the 2011 *International Conference on Information Systems* (Shanghai, China) for their valuable comments and suggestions. The authors also thank Tuan Quang Phan, Edoardo Airoldi, Wenjie Ping, Kar Tiong Khoo, and Cheryl Ong for assistance with the access and compilation of the data sets. All authors contributed equally to this research. This research is partially supported by the Singapore Ministry of Education, Project Grants R-253-000-071-112 and R-253-000-068-112.

References

- Adjei MT, Noble SM, Noble CH (2010) The influence of C2C communications in online brand communities on customer purchase behavior. *J. Acad. Marketing Sci.* 38(5):634–653.
- Akerlof GA (1970) The market for "Lemons": Quality uncertainty and the market mechanism. *Quart. J. Econom.* 84(3):488–500.
- Albuquerque P, Pavlidis P, Chatow U, Chen KY, Jamal Z (2012) Evaluating promotional activities in an online two-sided market of user-generated content. *Marketing Sci.* 31(3):406–432.
- Algesheimer R, Dholakia UM, Herrmann A (2005) The social influence of brand community: Evidence from European car clubs. *J. Marketing* 69(3):19–34.
- Amaldoss W, He C (2009) Direct-to-consumer advertising of prescription drugs: A strategic analysis. *Marketing Sci.* 28(3):472–487.
- Antweiler W, Frank MZ (2004) Is all that talk just noise? The information content of Internet stock message boards. *J. Finance* 59(3):1259–1294.
- Aral S, Walker D (2011) Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Sci.* 57(9):1623–1639.
- Arazy O, Kumar N, Shapira B (2010) A theory-driven design framework for social recommender systems. *J. Assoc. Inform. Systems* 11(9):455–490.
- Bagozzi RP, Dholakia UM (2006) Antecedents and purchase consequences of customer participation in small group brand communities. *Internat. J. Res. Marketing* 23(1):45–61.

¹⁹ We sincerely thank an anonymous reviewer for pointing out this limitation of our identification strategies.

²⁰ As a quick check of generalizability, we extract UGC and MGC from the Facebook fan page of another well-known apparel brand in the same country/market. We compare (using t-tests) weekly volumes of UGC and MGC of the FFS retailer's fan page to those of the other fan page but find no significant difference across these metrics.

- Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Quart. J. Econom.* 119(1):249–275.
- Bickart B, Schindler RM (2001) Internet forums as influential sources of consumer information. *J. Interactive Marketing* 15(3):31–40.
- Brown JJ, Reingen PH (1987) Social ties and word-of-mouth referral behavior. *J. Consumer Res.* 14(3):350–362.
- Burke M, Kraut R, Marlow C (2011) Social capital on Facebook: Differentiating uses and users. *ACM CHI Conf. Human Factors Comput. Systems* (ACM, New York), 571–580.
- Chen Y, Xie J (2008) Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Sci.* 54(3):477–491.
- Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(3):345–354.
- Chintagunta PK, Gopinath S, Venkataraman S (2010) The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Sci.* 29(5):944–957.
- Chu J, Chintagunta PK, Vilcassim NJ (2007) Assessing the economic value of distribution channels: An application to the personal computer industry. *J. Marketing Res.* 44(1):29–41.
- Clemons EK, Gao GG, Hitt LM (2006) When online reviews meet hyperdifferentiation: A study of the craft beer industry. *J. Management Inform. Systems* 23(2):149–171.
- Crosby LA, Stephens N (1987) Effects of relationship marketing on satisfaction, retention, and prices in the life insurance industry. *J. Marketing Res.* 24(4):404–411.
- Das SR, Chen MY (2007) Yahoo! for Amazon: Sentiment extraction from small talk on the Web. *Management Sci.* 53(9):1375–1388.
- Dellarocas C (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Sci.* 49(10):1407–1424.
- Dellarocas C (2006) Strategic manipulation of Internet opinion forums: Implications for consumers and firms. *Management Sci.* 52(10):1577–1593.
- Dhar V, Chang E (2009) Does chatter matter? The impact of user-generated content on music sales. *J. Interactive Marketing* 23(November):300–307.
- Duan W, Gu B, Whinston AB (2008) Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems* 45(4):1007–1016.
- Escalas JE (2007) Self-referencing and persuasion: Narrative transportation versus analytical elaboration. *J. Consumer Res.* 33(4):421–429.
- Facebook (2012) Facebook investor relations. Retrieved July 31, 2012, from <http://investor.fb.com/releasedetail.cfm?ReleaseID=695976>.
- Forman C, Ghose A, Wiesenfeld B (2008) Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inform. Systems Res.* 19(3):291–313.
- Ghose A, Ipeirotis PG (2011) Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Trans. Knowledge Data Engrg.* 23(10):1498–1512.
- Ghose A, Ipeirotis PG, Li B (2012) Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Sci.* 31(3):493–520.
- Gilly MC, Graham JL, Wolfinger MF, Yale LJ (1998) A dyadic study of interpersonal information search. *J. Acad. Marketing Sci.* 26(2):83–100.
- Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. *Marketing Sci.* 23(4):545–560.
- Godes D, Mayzlin D (2009) Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Sci.* 28(4):721–739.
- Goh KY, Hui KL, Png IPL (2011) Newspaper reports and consumer choice: Evidence from the do not call registry. *Management Sci.* 57(9):1640–1654.
- Grau J (2009) Social commerce on Facebook, Twitter and retail sites. Retrieved June 3, 2011, from http://www.emarketer.com/Reports/All/Emarketer_2000607.aspx.
- Gu B, Konana P, Rajagopalan B, Chen HWM (2007) Competition among virtual communities and user valuation: The case of investing-related communities. *Inform. Systems Res.* 18(1):68–85.
- Healey JS, Kassarian HH (1983) Advertising substantiation and advertiser response: A content analysis of magazine advertisements. *J. Marketing* 47(1):107–117.
- Heckman J (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Ann. Econom. Soc. Measurement* 5(4):475–492.
- Heckman J (1979) Sample selection bias as a specification error. *Econometrica* 47(1):153–161.
- Heckman J, Ichimura H, Todd P (1998) Matching as an econometric evaluation estimator. *Rev. Econom. Stud.* 65(2):261–294.
- Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.
- Kivetz R, Simonson I (2000) The effects of incomplete information on consumer choice. *J. Marketing Res.* 37(4):427–448.
- Kozinets RV (2002) The field behind the screen: Using netnography for marketing research in online communities. *J. Marketing Res.* 39(1):61–72.
- Liu Y (2006) Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing* 70(3):74–89.
- Luca M (2011) Reviews, reputation, and revenue: The case of Yelp.com. Harvard Business School NOM Unit Working Paper 12-016.
- Manchanda P, Xie Y, Youn N (2008) The role of targeted communication and contagion in product adoption. *Marketing Sci.* 27(6):961–976.
- Mayzlin D (2006) Promotional chat on the Internet. *Marketing Sci.* 25(2):155–163.
- McKenna KYA, Green AS, Gleason MEJ (2002) Relationship formation on the Internet: What's the big attraction? *J. Soc. Issues* 58(1):9–31.
- Mishra DP, Heide JB, Cort SG (1998) Information asymmetry and levels of agency relationships. *J. Marketing Res.* 35(3):277–295.
- Moe WW, Schweidel DA (2012) Online product opinions: Incidence, evaluation and evolution. *Marketing Sci.* 31(3):372–386.
- Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *J. Marketing Res.* 48(3):444–456.
- Muniz AM, O'Guinn TC (2001) Brand community. *J. Consumer Res.* 27(4):412–432.
- Nam S, Manchanda P, Chintagunta PK (2010) The effect of signal quality and contiguous word of mouth on customer acquisition for a video-on-demand service. *Marketing Sci.* 29(4):690–700.
- Narayanan S, Chintagunta PK, Miravete EJ (2007) The role of self selection, usage uncertainty and learning in the demand for local telephone service. *Quant. Marketing Econom.* 5(1):1–34.
- Nelson P (1970) Information and consumer behavior. *J. Political Econom.* 78(2):311–329.
- Obermiller C, Spangenberg ER (1998) Development of a scale to measure consumer skepticism toward advertising. *J. Consumer Psychol.* 7(2):159–186.
- Parks MR, Floyd K (1996) Making friends in cyberspace. *J. Comm.* 46(1):80–97.

- Pavlou PA, Dimoka A (2006) The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Inform. Systems Res.* 17(4):392–414.
- Porter CE, Donthu N (2008) Cultivating trust and harvesting value in virtual communities. *Management Sci.* 54(1):113–128.
- Ratchford BT (1982) Cost-benefit models for explaining consumer choice and information seeking behavior. *Management Sci.* 28(2):197–212.
- Rau LF, Jacobs PS, Zernik U (1989) Information extraction and text summarization using linguistic knowledge acquisition. *Inform. Processing Management* 25(4):419–428.
- Resnik A, Stern BL (1977) An analysis of information content in television advertising. *J. Marketing* 41(1):50–53.
- Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.
- Russo JE, Chaxel AS (2010) How persuasive messages can influence behavior without awareness. *J. Consumer Psychol.* 20(3):338–342.
- Schubert P, Ginsburg M (2000) Virtual communities of transaction: The role of personalization in electronic commerce. *Electronic Markets* 10(1):45–55.
- Smith NC, Cooper-Martin E (1997) Ethics and target marketing: The role of product harm and consumer vulnerability. *J. Marketing* 61(3):1–20.
- Sonnier GP, McAlister L, Rutz OJ (2011) A dynamic model of the effect of online communications on firm sales. *Marketing Sci.* 30(4):702–716.
- Tam KY, Ho SY (2005) Web personalization as a persuasion strategy: An elaboration likelihood model perspective. *Inform. Systems Res.* 16(3):271–291.
- Tirunillai S, Tellis G (2012) Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Sci.* 31(2):198–215.
- Trusov M, Bucklin RE, Pauwels K (2009) Effects of word-of-mouth versus traditional marketing: Findings from an Internet social networking site. *J. Marketing* 73(5):90–102.
- Tsai JY, Egelman S, Cranor L, Acquisti A (2011) The effect of online privacy information on purchasing behavior: An experimental study. *Inform. Systems Res.* 22(2):254–268.
- Tumarkin R, Whitelaw RF (2001) News or noise? Internet postings and stock prices. *Financial Analysts J.* 57(3):41–51.
- Von der Fehr NHM, Stevik K (1998) Persuasive advertising and product differentiation. *Southern Econom. J.* 65(1):113–126.
- Website-Monitoring (2010) Infographics. Retrieved September 7, 2011, from <http://www.website-monitoring.com/blog/category/infographics/>.
- Wu CC, Chen YJ, Wang CJ (2009) Is persuasive advertising always combative in a distribution channel? *Marketing Sci.* 28(6):1157–1163.
- Zhang X (2009) Retailers' multichannel and price advertising strategies. *Marketing Sci.* 28(6):1080–1094.
- Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Marketing* 74(2):133–148.